PLease do not answer if you are reading this

Paas, Leonard J.; Morren, Meike

published in
Marketing Letters
2018

DOI (link to publisher)
10.1007/s11002-018-9448-7

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Paas, L. J., & Morren, M. (2018). PLease do not answer if you are reading this: respondent attention in online panels. *Marketing Letters*, *29*(1), 13-21. https://doi.org/10.1007/s11002-018-9448-7

Mark Lett (2018) 29:13–21
https://doi.org/10.1007/s11002-018-9448-7

PLease do not answer if you are reading this: respondent attention in online panels

Leonard J. Paas[1] · **Meike Morren**[2]

Published online: 1 February 2018

Abstract This paper reports on the relevance of attention checks for online panels, e.g., M-Turk, SurveyMonkey, SmartSurvey, QualTrics. In two SmartSurvey studies approximately one third of the respondents failed a check that instructed them to skip the question. Attention-enhancing tools reduce this to approximately one fifth. The failure rate is not affected by replacing multiple-item scales with single-item measures. We find that failing the attention check relates to other attention indicators and that decreased attention levels often apply across the length of the survey. In support of relevance, our empirical findings show respondent inattentiveness systematically biases survey responses.

Keywords Respondent attention · Response bias · Instructional manipulation checks · Online panels · Single-item measurement

1 Introduction

Online panels such as M-Turk, SurveyMonkey, SmartSurvey, and QualTrics are often used for collecting survey data and for conducting experiments. Advantages include low prices and speedy data collection. However, some respondents are simultaneously subscribed to multiple panels (Comley 2005). Such "professional respondents" may dedicate the minimal cognitive effort required for providing plausible responses. This behavior could be more common in online surveys due to decreased personal contact (Johnson 2005) and anonymity (Meade and Craig 2012) which can result in low-quality data and faulty conclusions (Kaminska et al. 2010).

✉ Leonard J. Paas
l.j.paas@massey.ac.nz

[1] School of Communication, Journalism and Marketing, Massey University Albany, Private Bag 102904, North Shore Mail Centre, Auckland 0745, New Zealand

[2] Department of Marketing, VU University Amsterdam, Amsterdam, The Netherlands

Instructional manipulation checks (IMCs) are employed to assess respondent attention. In 2015, such checks were commonly applied in the four prominent marketing research and consumer behavior journals that we assessed. Three of the 51 *Marketing Letters* papers in 2015 reported IMCs. Out of these 51 papers, 31 reported lab experiments and/or primary collected survey data. Thus, in 9.7% (3/31*100%) of the papers in which IMCs were applicable such checks were indeed employed. This percentage is 8.1% for the 2015 issues of *Journal of Marketing Research*, 11.4% for *Journal of Consumer Psychology*, and 19.6% for *Journal of Consumer Research*.

Respondent inattention can be substantial. Oppenheimer et al. (2009) found up to 46% of lab experiment participants failed an IMC. Failure rates on IMCs vary in consecutive studies, e.g., 16.3 and 18.0% in the two survey studies reported by Emrich and Verhoef (2015) and 5.5% in Berman et al. (2015). Such differences may relate to survey length, respondent characteristics or the IMC employed (Meade and Craig 2012; Oppenheimer et al. 2009). Concerning the latter, some researchers used IMCs like those reported in Oppenheimer et al. (2009) whereas others created their own attention check: "What is the result of 7–5? If you're reading this question, please select 2 below" (2015): "How seriously do you take this study? Choose 'not at all seriously' if you read this question" (Barone et al. 2015).

The studies reported in the current paper embed the following IMC in a matrix of Likert-items: "This is a quality check. Please do not answer this question." This improves on previous IMC versions (Meade and Craig 2012; Oppenheimer Meyvis and Davidenko 2009) that often request a specific answer to be selected, e.g., "strongly agree." The requested category may be selected by attentive respondents but also coincidentally by respondents that are straight-lining the "strongly agree" category or others who are providing random answers.

This paper has three main contributions. First, study 1 reports findings on the validity of IMCs for online panels by assessing the relationship between IMC failure, on the one hand, and speeding, straight-lining and interest in the survey, on the other hand. Speeding and straight-lining are known indicators for respondent inattention (Oppenheimer et al. 2009; Zhang and Conrad 2013). Furthermore, we assess whether respondents failing an IMC early in the survey more often fail a later IMC. This would imply relatively consistent levels of inattention. Second, study 1 empirically analyzes whether IMC failure systematically biases responses, which is a more serious issue than random error. Third, we assess the effectiveness of tools for reducing inattentiveness. Some conditions in study 1 include an IMC at the beginning of the survey and/or an explicit warning in large red letters on a separate screen: "Note that there are checks in this survey to assess whether you are reading the survey questions properly." Moreover, study 2 assesses effects of question format on respondent attention by embedding an IMC amongst single-item measures instead of multiple-item scales. Next, we discuss the theoretical background to IMCs, followed by the two empirical studies and a discussion.

2 Theoretical background

Satisficing is the respondents' escape route to avoid thinking thoroughly about the survey (Krosnick 1991). Respondents may skip some steps in the response

process, i.e., gathering information in their mind to formulate a response, reflecting on which response category is most appropriate, etc. They may instead search for cues in questions enabling the selection of an answer that will be taken seriously by the interviewer (Krosnick 1999). Based on Krosnick's (1991) satisficing theory, Oppenheimer et al. (2009) propose that respondents dedicating limited cognitive resources will minimize their attention towards the survey, which may result in IMC failure.

Oppenheimer et al. (2009) found that respondents failing an IMC took less time to complete the experiment, i.e., speeding, and score lower on a need for cognition scale. However, they found no relationship with educational level. This could possibly be attributed to the relatively high level of homogeneity in their student sample. In an M-Turk sample Kapelner and Chandler (2010) find that the IMC pass rate is higher for women and increases with age. Again, no significant relationship is found with educational level or with need for cognition. Thus, the link between cognitive ability and IMC failure is uncertain. Instead, we propose that respondents who are interested in the survey's topic will dedicate more attention and fail the IMC less often. Previously reported findings indicate that interest in the topic enhances the likelihood of survey participation and decreases item non-response (Martin 1994).

We also suggest that satisficing, inattention, and the resulting IMC failure relate to response bias. The psychological literature distinguishes three main types of response bias: socially desirability, acquiescence, and extreme response bias (Van Herk, Poortinga, and Verhallen 2004). Social desirability bias occurs when the respondent carefully considers which answer is most acceptable in her/his societal context; satisficing respondents are unlikely to dedicate such attention to survey questions. Acquiescence is the tendency of respondents to agree with questions regardless of the content. Weijters et al. (2013) found that IMC failure does not relate strongly to acquiescence bias when measured by reversed questions. The use of balanced scales that consist of reversed and positively formulated questions is the gold standard in distinguishing between acquiescence response bias and an actual positive attitude (Billiet and McClendon 2000). Extreme response bias occurs when participants respond more extremely than one would expect when considering their attitudes (Van Herk et al. 2004). Extreme responding is more likely to get noticed, which contradicts the employment of satisficing strategies.

Inattentive respondents may instead choose a response that can easily be argued for, such as the neutral midpoint category. When the selection of the midpoint category is repeated throughout the survey and occurs regardless of question content this response behavior is called straight-lining (Holbrook, Green and Krosnick 2003). In previously reported empirical studies straight-lining has been associated with low attention levels (Zhang and Conrad 2013, Malhotra 2008). We propose that straight-lining the midpoint can infer systematic bias, i.e., means and standard deviations no longer reflect true values on the measured variables. For example, if the inattentive respondent is straight-lining the Likert-scale midpoint on questions to which other respondents tend to agree strongly this response behavior will result in lower mean scores. Furthermore, straight-lining may take effect when considering many questions that treat diverse topics implying lower standard deviations (Zhang and Conrad 2013).

Study 1 tests consistency of the IMC with the other inattention indicators. This study also tests whether IMC failure is consistent across the length of the survey, relates to

systematic response bias and assesses IMC failure reductions resulting from attention-enhancing tools. Study 2 tests whether single-item measures reduce IMC failure.

3 Study 1: IMCS and enhancing attention in an online panel

3.1 Sample and method

Study 1 includes 418 SmartSurvey respondents who completed the survey after removing 13 non-Dutch respondents; surveys were in the Dutch language. Respondents are aged 14 to 79 years ($m = 46.8$) and have different educational levels and 51.0% are female. They received a €0.50 payment, which is equivalent to approximately US$0.60 and similar to payments in other online panels. Some experimental conditions embedded an IMC earlier in the survey, after the first seven questions, and/or a warning about attention checks, after another five questions, leading to the 2-by-2-design in Table 1. The earlier IMC is intended as an attention-enhancing tool similar to the warning. The later IMC is embedded after 39 questions and aims to measure respondent attention.

Respondents in all conditions answered 41 survey questions of which 36 are seven-point Likert-scales (excluding the embedded IMCs). They were asked about characteristics such as age, gender and reacted to various attitude questions. Next, they were exposed to an advertisement, and we measured ad attitude using Holbrook and Batra's (1987) four-item Likert-format scale; Putrevu and Lord's (1994) five-item Likert-format scale measured brand attitude, and their three-item Likert-format scale measured purchase intentions. As in previous studies these scales have sufficient psychometric properties in our data.

We measured time spent on the survey (speeding) and counted the number of questions to which a respondent selected the same answer-category (straight-lining). The respondent's interest in the survey is measured using a 7-point Likert-scale embedded at the end of the survey. Concerning the straight-lining variable, consider a respondent selecting "strongly agree" on 20 Likert-scale questions and other categories on 16 Likert-scales; the straight-line variable takes on the value 20 for this specific individual. If another respondent selected the same category on 30 Likert-scales, instead of 20, the straight-lining score will be 30. The most extreme respondent answers the same category on all 36 Likert-scales. However, respondents may expect such

Table 1 Experimental design and outcomes (study 1)

	No warning	Warning
No early IMC	Condition 1	Condition 2
	• IMC failure 35.2% ($n = 105$)	• IMC failure 19.6% ($n = 102$)
	• Completion 84.7% ($n = 124$)	• Completion 75.6% ($n = 135$)
Early IMC	Condition 3	Condition 4
	• IMC failure 20.4% ($n = 103$)	• IMC failure 19.4% ($n = 108$)
	• Completion 90.3% ($n = 114$)	• Completion 88.5% ($n = 122$)

extremes to get noticed, especially in the warning condition. Thus, they may deviate from the midpoint when answering some questions (Zhang and Conrad 2013).

3.2 Results

Validity of IMC failure in online panels The logistic regression analysis results reported in Table 2 show that IMC failure is consistent with other inattention indicators. The dependent variable takes on one of two possible values for each respondent: failed the IMC or passed. The model includes the following independent variables: time spent on the survey (in minutes), straight-lining and interest in the survey. We assume no causality in this analysis. Response times are capped at 11 min to avoid overly large effects of extreme values. Figure 1 shows that short and very long response times relate to higher IMC failure probabilities. Thus, we add a quadratic term for response time. Table 2 shows that the Nagelkerke r^2 equals 0.27 ($p < 0.01$); intermediate response times, less straight-lining and more interest in the survey relate to less IMC failure.

Validity of the IMC is further supported by consistency in failing the early and later IMC. Amongst the 211 respondents in conditions three and four of Table 1, only 5.5% of the respondents not failing the early IMC do fail the IMC at the end. However, 75.6% of the respondents failing the early IMC also fail the IMC embedded at the end of the survey ($d.f. = 1$, $\chi^2 = 99.04$, $p < 0.01$).

IMC failure and response bias We find that IMC failure is related to systematic response bias, which is a more serious concern than random error. The mean score on ad attitude for IMC-failers ($n = 99$), $m = 4.30$, does not differ significantly from the mean for others ($n = 319$), $m = 4.15$, $d.f. = 416$, $t = 1.45$, $p > 0.10$. However, there are significant differences in brand attitude scores, i.e., $m = 3.98$ for IMC-failers versus $m = 3.47$ for others, $d.f. = 416$, $t = 5.65$, $p < 0.01$. Furthermore, for purchase intentions, Levene's test for equality of variances is significant, $F = 5.10$, $p < 0.05$, implying that the variation on this variable differs significantly between IMC-failers and those respondents passing. The t test with equal variances not assumed shows that the mean on purchase intentions for IMC-failers, $m = 4.06$, differs significantly from other respondents, $m = 3.68$, $d.f. = 162.29$, $t = 3.37$, $p < 0.01$.

We test whether IMC failure affects responses to survey questions beyond speeding using three two-way ANOVAs with ad attitude, brand attitude, respectively, purchase intention scores as dependents. The two independent variables in each of these three

Table 2 Speeding, straight-lining interest in the topic and failing the IMC (study 1)*

Variable	Beta	Exp(Beta)	p value
Intercept	− 2.20	8.98	0.46
Minutes	− 0.67	1.79	0.00
Minutes^2	0.05	0.96	0.00
Interest in topic	− 0.26	1.30	0.01
# same category	0.15	0.86	0.00

* The dependent variable is IMC failure. Omnibus test of model coefficients $d.f. = 4$, $\chi^2 = 81.56$, $p < 0.01$, and Nagelkerke $r^2 = 0.27$

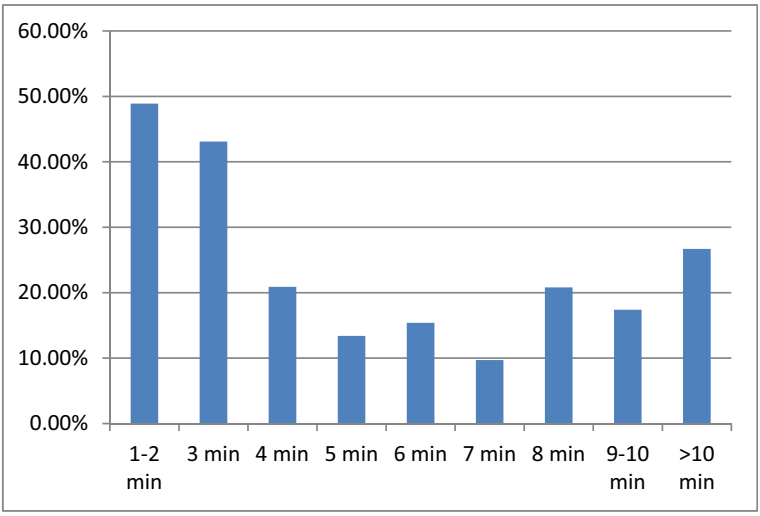

Fig. 1 IMC failure rate and time spent on survey (study 1)*. * Minutes are rounded, e.g., 3 min indicates the respondent dedicated between 2 min and 30 s and 3 min and 30 s to the survey

ANOVAs are a binary speeding indication (yes/no) and the IMC variable (failed/passed). Respondents dedicating less than 0.3 s on average for reading each word in the survey are defined as speeders (Zhang and Conrad, 2013). Speeding significantly affects ad attitude scores ($d.f. = 1$, $F = 5.44$, $p < 0.05$), brand attitude ($d.f. = 1$, $F = 11.74$, $p < 0.01$), and purchase intentions ($d.f. = 1$, $F = 9.54$, $p < 0.01$). Effects of IMC failure on brand attitude and purchase intentions scores are also significant, i.e., $d.f. = 1$, $F = 12.15$, $p < 0.01$, respectively, $d.f. = 1$, $F = 6.09$, $p < 0.05$, but not for ad attitude, $d.f. = 1$, $F = 0.56$, $p > 0.10$. Thus, IMC failure enhances detection of inattentive respondents when used in combination with speeding. This is also supported by the finding that only 30% of the respondents failing the IMC ($n = 99$) are also defined as speeders.

Effects of attention-enhancing tools We applied logistic regression to analyze the effects of attention-enhancing tools on IMC failure at the end of the survey. Following the experimental design in Table 1, we include two binary independents: (1) the respondent received an IMC at the beginning of the survey (no/yes); (2) exposure to the warning (no/yes). The interaction term between these two variables is also modeled. Table 3 reports

Table 3 Effects of attention-enhancing tools on failing the IMC (study 1)*

Variable	Beta	Exp(Beta)	*p* value
Intercept	−0.61	1.84	0.00
Early IMC	−0.75	2.12	0.02
Warning	−0.80	2.23	0.01
Warning * Early IMC	0.74	0.48	0.12

* The dependent variable is IMC failure. Omnibus test of model coefficients $d.f. = 3$, $\chi^2 = 9.82$, $p < 0.05$; Nagelkerke $r^2 = 0.04$. We assessed significance using bootstrapping to accommodate for the presence of an interaction term (Echambadi and Hess 2007)

that the early IMC and the warning both reduce the probability of IMC failure at the survey's end. However, the interaction effect is insignificant. Although Table 3 reports a modest Nagelkerke r^2, the difference between failure rates in condition 1, without the attention-enhancing tools, and conditions 2 to 4, with the warning and/or early IMC, is approximately one third versus one fifth, see Table 1.

We find no indications for adverse effects of the attention-enhancing tools. The binary variable survey completion (yes/no) was included as the dependent variable in a logistic regression analysis with the following two independents: IMC at the beginning (no/yes) and warning (no/yes). The interaction term between these two variables is also modeled. We find no significant effects on survey completion, although the effect of the warning is negative and close to significance ($p = 0.07$). Furthermore, three separate two-way ANOVAs show attention-enhancing tools do not significantly affect the means on the ad attitude, brand attitude, and purchase intentions scales (all $p > 0.10$).

4 Study 2: IMC failure in single-item measurement

Besides attention-enhancing tools questionnaire format may also relate to IMC failure. Bergkvist and Rossiter (2007) proposed single-item measures for ad- and brand attitude. A similar argument is applicable for purchase intentions. Although a recent paper criticizes single-item scales (Kamakura 2015), we propose that respondents may dedicate more attention to such measures resulting from reduced repetition (Rossiter 2002), which relates to the purpose of IMCs.

Study 2 replaces the multiple-item scales for measuring ad attitude, brand attitude, and purchase intentions with the single item having the highest factor loading in the Study 1 data. Other than that, the study 2 procedure is the same as for study 1. As in condition 1 of study 1, we did not use attention-enhancing instruments and include the same IMC near the survey's end, amongst the single-item measures.

We analyzed data from 107 Dutch SmartSurvey respondents, after removing four non-Dutch respondents. Demographics of study 2 respondents are comparable to study 1. Thus, the only difference between study 2 and condition 1 of study 1 concerns the use of single-item scales. We find that 32.1% of the study 2 respondents fail the IMC. This does not differ significantly from the failure rate of 35.2% found in condition 1 of study 1 ($z = 0.48$, $p > 0.10$).

5 Discussion

Satisficing respondents answer questions without dedicating sufficient attention (Meade and Craig 2012; Oppenheimer et al. 2009; Zhang and Conrad 2013). The study 1 results show that IMC-failers often dedicate less time to the survey. This result is consistent with findings that are reported by Oppenheimer et al. (2009); the non-linear relationship between response time and IMC failure is a novel finding. Furthermore, we find that respondents failing the IMC more often straight-line and also are less interested in the survey's topic. Study 1 also shows that IMC failure applies across the length of the survey, i.e., many respondents in conditions 3 and 4 of Table 1 fail both

the first and second embedded IMC. Furthermore, we find that reduced respondent attention levels, as measured by IMCs, systematically bias responses even after taking speeding into consideration. These findings support the relevance of IMCs in online panels. Further research may assess generalizability of our novel findings, i.e., the non-linear relationship between time spent on the survey and IMC failure, the reported relationship between IMC failure early in the survey and later IMC failure and the finding that IMC failure relates to systematic response bias.

Relatively simple attention-enhancing instruments reduced IMC failure in study 1 from approximately one third to one fifth. The early IMC and the warning did not affect survey completion rates and did not result in biased survey responses, which further supports their applicability. Attention-enhancing tools cannot be replaced by employing single-item measures; study 2 results show that such measures do not reduce IMC failure.

As a limitation, we find that the attention-enhancing instruments do not fully mitigate respondent inattention. Future research could assess whether higher payments can further enhance attention. However, it may be impossible to motivate some respondents, such as the study 1 participants failing both IMCs in conditions 3 and 4 of study 1. Online panel firms could aim to detect such individuals and then warn them about their response behavior or remove them from the panel. Future studies may also analyze other data sources, e.g., face-to-face pencil-and-paper surveys, lab experiments. Another avenue for future research concerns assessing potential caveats of IMCs, e.g., perceived reductions in anonymity or respondent irritation.

Acknowledgements The authors are grateful for feedback provided by Harald van Heerde, the marketing group at the Radboud University and the School of Communication, Journalism and Marketing at Massey University. We also express our gratitude to Siara Khan for assessing the use of attention checks in the marketing literature.

References

Barone, M. J., Lyle, K. B., & Winterich, K. P. (2015). When deal depth doesn't matter: how handedness consistency influences consumer response behavior to horizontal versus vertical price comparisons. *Marketing Letters, 26*(2), 213–223.

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same construct. *Journal of Marketing Research, 44*(2), 175–184.

Berman, J. Z., Levine, E. E., Brach, A., & Small, D. A. (2015). The braggart's dilemma: on the social rewards and penalties of advertising prosocial behaviour. *Journal of Marketing Research, 52*(1), 90–104.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*(4), 608–628.

Comley, P. (2005). Understanding the online panelist. *ESOMAR*. http://www.websm.org/db/12/11356/Bibliography/Understanding_the_online_panellist.

Echambadi, R., & Hess, J. P. (2007). Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science, 26*(3), 438–445.

Emrich, O., & Verhoef, P. C. (2015). The impact of a homogenous versus a prototypical web design on online retail patronage for multichannel providers. *International Journal of Research in Marketing, 32*(4), 363–374.

Holbrook, M. B., & Batra, R. (1987). Assessing the role of emotions as mediators of consumer responses to advertising. *Journal of Consumer Research, 14*(4), 404–420.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires. Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*(1), 79–125.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*(1), 103–129.

Kamakura, W. A. (2015). Measure twice and cut once: the carpenter's rule still applies. *Marketing Letters, 26*(3), 237–243.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly, 74*(5), 956–984.

Kapelner, A., & Chandler, D. (2010). *Preventing satisficing in online surveys: a "Kapcha" to ensure higher quality data*. San Francisco: Paper presented at the Crowd Conference.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213–236.

Krosnick, J. A. (1999). Survey research. *Annual Review Psychology, 50*, 537–567.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly, 72*(5), 914–934.

Martin, C. L. (1994). The impact of topic interest on mail survey response behavior. *Journal of the Marketing Research Society, 36*(4), 327–338.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–972.

Putrevu, S., & Lord, K. (1994). Comparative and non-comparative advertising: Attitudinal effects under cognitive and affective involvement conditions. *Journal of Advertising, 23*(2), 77–91.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal for Research in Marketing, 19*(4), 305–335.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*(3), 346–360.

Weijters, B., Baumgartner, H. and Schillewaert (2013). Reversed item bias: an integrative model. Psychological Methods, 18(3), 320–334.

Zhang, C., & Conrad, F. G. (2013). Speeding in web suveys: the tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135.